



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

Address: COMMISSIONER FOR PATENTS

P.O. Box 1450

Alexandria, Virginia 22313-1450

www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/662,602	09/15/2003	Chang-Ning Huang	M61.12-0514	2369
27366 7590 05/19/2009 WESTMAN CHAMPLIN (MICROSOFT CORPORATION) SUITE 1400 900 SECOND AVENUE SOUTH MINNEAPOLIS, MN 55402				
EXAMINER				
SAINT CYR, LEONARD				
ART UNIT		PAPER NUMBER		
2626				
MAIL DATE		DELIVERY MODE		
05/19/2009		PAPER		

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary

Application No.

10/662,602

Applicant(s)

HUANG ET AL.

Examiner

LEONARD SAINT CYR

Art Unit

2626

Period for Reply -- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 03 March 2009.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 47 - 66 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 47 - 66 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 09/15/03 is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO-8508)
Paper No(s)/Mail Date _____
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date _____
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: _____

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

1. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 03/03/09 has been entered.

Claim Rejections - 35 USC § 102

2. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

3. Claims 47 - 52, 57 – 59 are rejected under 35 U.S.C. 102(b) as being anticipated by Palmer et al., (Chinese Word Segmentation and Information Retrieval; 1997).

As per claims 47, and 57, Palmer et al., teach a computer-implemented method for evaluating a word segmentation language model, comprising:

building the word segmentation language model ("segmentation based on character bigrams"; page 175, col.2, paragraph 1; page 177, col.2, paragraph 5);

utilizing a computer processor that is a functional component of the computer to apply the language model to a test corpus of unsegmented text different from the annotated corpus so as to provide an output indicative of words in the test corpus and a word type indication for each word, the word type indication being any one of a plurality of word type indications ("**refinement steps attempt to recognize idiomatic expressions, derived words, Chinese person names, and foreign proper names**" those are considered as word type indications; pages 177, col.2, paragraph 3);

utilizing the processor to compare the word type indication for each word in the output of the language model with predefined word type indications of words of the test corpus; and utilizing the processor to generate a quantitative value that represents a level of precision with which word type indications were applied in the output indicative of words in the test corpus, wherein generating comprises generating based on a comparison of the word type indication for words in the output to the predefined word type indications ("precision is defined as the percentage of identified words which are also in the same positions in the hand-segmented text... **refinement steps attempt to recognize idiomatic expressions, derived words, Chinese person names, and foreign proper names**, determine the contribution of each of these steps to the segmentation accuracy"; page 176, col.2, paragraph 5; pages 177, col.2, paragraph 3).

As per claim 48, Palmer et al., further disclose building the word segmentation language model comprises building the word segmentation language model based on an annotated corpus ("segmentation based on character bigrams"; page 175, col.2,

paragraph 1; page 177, col.2, paragraph 5).

As per claims 49, and 58, Palmer et al., further disclose generating the quantitative value further comprises generating a quantitative value based on a comparison of word type indications of words in the output that match predefined word type indications assigned to the same words in the test corpus indications ("precision is defined as the percentage of identified words which are also in the same positions in the hand-segmented text"; page 176, col.2, paragraph 5).

As per claim 50, Palmer et al., further disclose generating the quantitative value comprises generating a quantitative value that is indicative of how frequently a word type indication in the output matches a corresponding predefined word type indication in the test corpus ("use **frequency-based phrase building, that is, segmentation based** on character n-gram occurrences in the collection"; page 177, col.2, paragraph 5).

As per claim 51, Palmer et al., further disclose generating the quantitative value comprises generating a quantitative value that is indicative of how frequently a word type indication, assigned to a word in the output, matches a predefined word type indication assigned to a same word in the test corpus ("use **frequency-based phrase building, that is, segmentation based** on character n-gram occurrences in the collection"; page 177, col.2, paragraph 5).

As per claims 52, and 59, Palmer et al., further disclose generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which person name word type indications were applied in the output ("Chinese person names"; page 177, col.2, paragraph 3).

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. Claims 53, 54, 60, and 61 are rejected under 35 U.S.C. 103(a) as being unpatentable over Palmer et al., (Chinese Word Segmentation and Information Retrieval; 1997) in view of Guo et al., (US PAP 2002/0052901).

As per claims 53, 54, 60, and 61, Palmer et al., do not specifically teach that location name type indications were applied in the output; organization name word type indications were applied in the output.

Guo et al., teach that for Chinese language, the morphologic process includes the steps of: (1) segmenting sentences into words according to the system dictionary and the user-defined dictionaries; (2) identifying proper names (currently including person names, **place names** and person titles), **domain terms**, numbers, measure words, and date expressions (paragraph 29).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made use place names and domain terms in precision scores as taught by Guo et al., in Palmer et al., because that would contribute to the segmentation accuracy (Palmer et al., page 177, col.2, paragraph 3).

6. Claims 55, 56, and 62 are rejected under 35 U.S.C. 103(a) as being unpatentable over Palmer et al., (Chinese Word Segmentation and Information Retrieval; 1997) in view of Haizhou et al., (Chinese Word Segmentation, 1998).

As per claims 55, 56, and 62, Palmer et al., do not specifically teach that overlapping ambiguous string word type indications were applied in the output; organization name word type indications were applied in the output; covering ambiguous string word type indications were applied in the output.

Haizhou et al., teach As indicated by Liang[2], there are two cases of unexpected segmentation. One is overlapping ambiguity where a character could go either way to form two words, such as in example 1. Another is composition ambiguity where the subsegmentation is possible:...One can find that , and all are possible word entries, thus both results are valid based on lexicon entries (page 215, section 3.2, paragraph 3).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made use overlapping ambiguous string word and covering ambiguous string word in precision scores as taught by Haizhou et al., in Palmer et al., because that would contribute to the segmentation accuracy (Palmer et al., page 177, col.2, paragraph 3).

7. Claims 63 64, and 66 are rejected under 35 U.S.C. 103(a) as being unpatentable over Palmer et al., (Chinese Word Segmentation and Information Retrieval; 1997) in view of Dien et al., (Vietnamese Word Segmentation, 2001).

As per claim 63, Palmer et al., teach a computer-implemented method for performing word segmentation, the method comprising:

receiving an input of text; utilizing a computer processor that is a functional component of the computer to apply a language model so as to determine a segmentation of the text ("segmentation based on character bigrams"; page 175, col.2, paragraph 1; page 177, col.2, paragraph 5);

identifying a morphologically derived word within the text; and providing an output that includes the segmentation of the text ("derived words...to the segmentation accuracy"; page 177, col.2, paragraph 3).

However, Palmer et al., do not specifically teach an indication of a combination of parts that form the morphologically derived word.

Dien et al., teach that those are the words morphologically derived (R.Sproat,1996), e.g.: coá gaéng (attempt, v) sǎi coá gaéng (attempt, n) hieăi ñăi (modern, a) hieăi ñăi hoà (modernize) chuû tòch (president,n) phoû chuû tòch (vice-president) [section 4.1.5].

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made use indication of a combination of parts that form derived

words as taught by Dien et al., in Palmer et al., because that would contribute to the segmentation accuracy (Palmer et al., page 177, col.2, paragraph 3).

As per claim 64, Palmer et al., further disclose the output also includes an indication of a named entity detected within the text ("Chinese person names"; page 177, col.2, paragraph 3).

As per claim 66, Palmer et al., further disclose the output also includes an indication a part of speech represented by the combination of parts [section 4.1.5].

8. Claim 65 is rejected under 35 U.S.C. 103(a) as being unpatentable over Palmer et al., (Chinese Word Segmentation and Information Retrieval; 1997) in view of Dien et al., (Vietnamese Word Segmentation, 2001), and further in view of Guo et al., (US PAP 2002/0052901).

As per claim 65, Palmer et al., in view of Dien et al., do not specifically teach the output also includes an indication of a factoid detected within the text.

Guo et al., teach that for Chinese language, the morphologic process includes the steps of: (1) segmenting sentences into words according to the system dictionary and the user-defined dictionaries; (2) identifying proper names (currently including person names, place names and person titles), domain terms, **numbers, measure words, and date expressions** (paragraph 29).

Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to identify factoids as taught by Guo et al., in Palmer et al., because that would contribute to the segmentation accuracy (Palmer et al., page 177, col.2, paragraph 3).

Conclusion

9. Any inquiry concerning this communication or earlier communications from the examiner should be directed to LEONARD SAINT CYR whose telephone number is (571) 272-4247. The examiner can normally be reached on Mon- Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571) 272-7602. The fax phone number for the organization where this application or proceeding is assigned is (571)-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Application/Control Number: 10/662,602

Page 10

Art Unit: 2626

/Richmond Dorvil/

Supervisory Patent Examiner, Art Unit 2626